PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Natalizumab versus fingolimod for patients with active relapsing-remitting multiple sclerosis: results from REVEAL, a prospective, randomised head-to-head study
AUTHORS	Butzkueven, Helmut; Licata, Stephanie; Jeffery, Douglas; Arnold, Douglas; Filippi, Massimo; Geurts, Jeroen; Santra, Sourav; Campbell, Nolan; Ho, Pei-Ran

VERSION 1 – REVIEW

REVIEWER	Dr Özgür Yaldizli University Hospital Basel Neurologic clinic and policlinic
REVIEW RETURNED	26-Apr-2020

GENERAL COMMENTS	This study is a head-to-head RCT to compare the efficacy of Natalizumab vs. Fingolimod in RRMS. The study was conducted in 9 countries (43 sites). In total, 108 patients were randomized 1:1 to natalizumab or fingolimod in a standard dose. The study was designed to include 540 patients. However, after one year of recruitment only 111 patients had been enrolled. Due to slow recruitment of patients, the sponsor terminated the study. Inclusion criteria were patients between 18-60 years with active RRMS not previously treated with natalizumab, fingolimod or any other immunosuppressants. Patients had to have at least one new T1 gad+ lesion within the 6 months prior to screening or at least 2 new T2 lesions on brain MRI within the 6 months prior to screening compared with a previous scans 18 months before screening. The EDSS had to be <6.0. The patients were allowed to be on GA or an IFN beta; moreover, they had to have at least nine T2 hyperintense lesions on brain MRI and at least one relapse on therapy within 6 months prior to the screening. The inclusion criteria for treatment-naïve MS patients or patients who were treated for <6 months with GA or IFN beta were stricter. They had to have at least 2 relapses within the 12 months prior to screening and these relapses had to have been disabling. The two treatment arms were well balanced. The IIT analysis revealed a 63% lower mean number of T1 Gad+ lesions in the natalizumab groups vs. the fingolimod group at 4 weeks which did not reach statistical significance. However, at 12 and 24 weeks the difference between the groups regarding Gad+ lesions reached statistical significance. Over 6 months, patients on natalizumab were less likely to develop 2 or more new T1 Gad+ lesions. Moreover, natalizumab treated patients were less likely to experience a relapse. The risk for a relapse on study was 1.9% on natalizumab vs. 22% on fingolimod during the first 6 months. The on-treatment ARR was about 80% lower in natalizumab vs. fingolimod treated patients. The AE rate was marginally higher for fingolimod than natalizumab. Two

	serious AEs were reported for fingolimod including a second-degree AV block and a migraine with aura. The authors of the study concluded that natalizumab is more efficacious than fingolimod in reducing relapses and T1 Gd+ lesion accumulation in patients with active disease over 6 months. This manuscript is excellently written. As the authors stated, there are already several studies in real-world that have compared fingolimod and natalizumab. They uniformly showed stronger suppression of disease activity for natalizumab vs. fingolimod in RRMS. The new aspect of this study is that it is a RCT. Unfortunately, this study failed to reach initially planned number of patients. Nevertheless, parts of the analysis, especially the analysis of Gadolinium enhancing lesions between the groups can be used as a surrogate for the comparison of suppression of disease activity in active RRMS. Here, the study shows what the reader would expect, namely that natalizumab is stronger anti-inflammatory than fingolimod. Moreover, the study provides data about safety and tolerability. Of course, this data is not new; however, worth reporting. I recommend to accept this manuscript. The only minor comment from my side is that the authors should again review the literature and reference all existing literature comparing fingolimod vs. natalizumab in RRMS; the referenced studies are not the only studies that compare the effectiveness of natalizumab vs. fingolimod. Alternatively, a review including all existing literature to the comparison between natalizumab vs. fingolimod would be sufficient.

REVIEWER	Damiano Baroncini Multiple Sclerosis Centre Gallarate Hospital ASST Valle Olona Gallarate (VA), Italy In the last three years I received travel grants for attending congresses and meetings, speaking and scientific publications honoraria from Novartis, Roche, Genzyme, Merck, Biogen and Almirall.
REVIEW RETURNED	09-May-2020

GENERAL COMMENTS	The authors presented data of a randomized, single blind, controlled, head-to-head study comparing natalizumab Vs fingolimod in patients with active RRMS. The study was early closed by the sponsor (Biogen) due to a slow enrolment. Primary endpoint could not be analysed then, but secondary endpoints analyses were available. Over 24 and 36 weeks natalizumab performed better than fingolimod in reducing new T1 Gd+ lesion formation and relapses respectively. No difference was found in the other MRI endpoints (new/enlarging T2 lesions, T2 lesion volumes changes). 50% were naïve patients. A subanalysis found a significant trend for a faster reduction of new T1 Gd+ lesion formation in natalizumab group in the first 8 weeks, which became significant after 12 weeks. This research confirms again the higher power of natalizumab in controlling inflammatory disease activity with respect to fingolimod. The main limitations of the study are small sample size and short duration of follow-up. However, analyses of MRI endpoints catch the speed of the effect, which was higher for natalizumab.

	The paper is well written and clear. However, in order to further improve the work, there are some issues to address. Statistics 1) Methods, page 10, row 13: Wilcoxon rank-sum test has to be used for dependent sample comparison (i.e. repeated measure within the same group of subjects). Here there is no such analysis, therefore it should be used the Mann-Whitney test for comparison of number of new/newly enlarging T2 lesions and changes in lesion volume between natalizumab and fingolimod groups (two independent samples). 2) Table 3: the negative binomial model for ARR comparison should also be adjusted for variable “years since first symptoms”, which is disease duration, as was done for Cox model for T1 Gd+ lesions. Others (style, presentation) 3) Introduction, row 22: I would extend this part by citing other important studies, explaining that in most natalizumab proved to be superior to fingolimod (Carruthers 2014; Kalincik 2015; Barbin 2016; Baroncini 2016; Prosperini 2017), while in some it did not (Braune 2013; Koch-Henriksen 2016). 4) Introduction, row 30 to 42: I would erase all the sentence (“Although…over the duration of the study”), because it is clearly explained in methods section. 5) Methods, page 8, row 15: I would add “…the sponsor (BIOGEN)…” in the sentence 6) Table 1: standard deviation (SD) should have the same number of decimals of the mean (I would keep only one decimal) 7) Table 3: I would keep only one decimal for all values (both for means and SD), except for HR and p values. 8) Table 3: It should be correct with the results of Mann-Whitney test instead of Wilcoxon test 9) Discussion, page 16, row 10: the same issues as for the introduction in point 3)

REVIEWER	Susan Ball University of Exeter, UK
REVIEW RETURNED	26-May-2020

GENERAL COMMENTS	The abstract is clear and easy to understand. The authors describe the intention-to-treat population as those who were randomised and given at least 1 dose of study drug and who provided any efficacy assessments. Would this be better described as a modified intention-to-treat population? The authors have only been able to conduct analyses of secondary endpoints. Are these analyses based on a smaller sample of participants and/or a shorter follow-up period than originally planned for assessing these endpoints? Were these analyses always intended to be exploratory? The statistical methods used seem appropriate and well described. When reporting the results from Negative Binomial regression, would it be possible to report a rate in each trial arm and a rate ratio, with confidence interval? Confidence intervals for differences in rates between the two arms may be more informative than p values. Table 3 - could a rate ratio be reported in the 'HR (95% CI)' column for ARR on study?

	Figure 2 - Kaplan Meier plot - check the labeling of vertical axis - seems to be on the wrong scale, compared to the text '22.3%' and '1.9%' on the plot.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Dr Özgür Yaldizli
Institution and Country: University Hospital Basel, Neurologic clinic and policlinic, Switzerland
Please state any competing interests or state 'None declared': None declared

This study is a head-to-head RCT to compare the efficacy of Natalizumab vs. Fingolimod in RRMS. The study was conducted in 9 countries (43 sites). In total, 108 patients were randomized 1:1 to natalizumab or fingolimod in a standard dose. The study was designed to include 540 patients. However, after one year of recruitment only 111 patients had been enrolled. Due to slow recruitment of patients, the sponsor terminated the study. Inclusion criteria were patients between 18-60 years with active RRMS not previously treated with natalizumab, fingolimod or any other immunosuppressants. Patients had to have at least one new T1 gad+ lesion within the 6 months prior to screening or at least 2 new T2 lesions on brain MRI within the 6 months prior to screening compared with a previous scans 18 months before screening. The EDSS had to be <6.0. The patients were allowed to be on GA or an IFN beta; moreover, they had to have at least nine T2 hyperintense lesions on brain MRI and at least one relapse on therapy within 6 months prior to the screening. The inclusion criteria for treatment-naïve MS patients or patients who were treated for <6 months with GA or IFN beta were stricter. They had to have at least 2 relapses within the 12 months prior to screening and these relapses had to have been disabling. The two treatment arms were well balanced. The IIT analysis revealed a 63% lower mean number of T1 Gad+ lesions in the natalizumab groups vs. the fingolimod group at 4 weeks which did not reach statistical significance. However, at 12 and 24 weeks the difference between the groups regarding Gad+ lesions reached statistical significance. Over 6 months, patients on natalizumab were less likely to develop 2 or more new T1 Gad+ lesions. Moreover, natalizumab treated patients were less likely to experience a relapse. The risk for a relapse on study was 1.9% on natalizumab vs. 22% on fingolimod during the first 6 months. The on-treatment ARR was about 80% lower in natalizumab vs. fingolimod treated patients. The AE rate was marginally higher for fingolimod than natalizumab. Two serious AEs were reported for fingolimod including a second-degree AV block and a migraine with aura. The authors of the study concluded that natalizumab is more efficacious than fingolimod in reducing relapses and T1 Gd+ lesion accumulation in patients with active disease over 6 months.
This manuscript is excellently written. As the authors stated, there are already several studies in real-world that have compared fingolimod and natalizumab. They uniformly showed stronger suppression of disease activity for natalizumab vs. fingolimod in RRMS. The new aspect of this study is that it is a RCT. Unfortunately, this study failed to reach initially planned number of patients. Nevertheless, parts of the analysis, especially the analysis of Gadolinium enhancing lesions between the groups can be used as a surrogate for the comparison of suppression of disease activity in active RRMS. Here, the study shows what the reader would expect, namely that natalizumab is stronger anti-inflammatory than fingolimod. Moreover, the study provides data about safety and tolerability. Of course, this data is not new; however, worth reporting. I recommend to accept this manuscript.

The only minor comment from my side is that the authors should again review the literature and reference all existing literature comparing fingolimod vs. natalizumab in RRMS; the referenced studies are not the only studies that compare the effectiveness of natalizumab vs. fingolimod. Alternatively, a review including all existing literature to the comparison between natalizumab vs. fingolimod would be sufficient.

Author response: The Introduction has been revised to include a more comprehensive review of the literature comparing natalizumab and fingolimod (page 5, paragraph 1).

Reviewer: 2
Reviewer Name: Damiano Baroncini
Institution and Country: Multiple Sclerosis Centre, Gallarate Hospital, ASST Valle Olona Gallarate (VA), Italy Please state any competing interests or state 'None declared': In the last three years I received travel grants for attending congresses and meetings, speaking and scientific publications honoraria from Novartis, Roche, Genzyme, Merck, Biogen and Almirall.

The authors presented data of a randomized, single blind, controlled, head-to-head study comparing natalizumab Vs fingolimod in patients with active RRMS. The study was early closed by the sponsor (Biogen) due to a slow enrolment.
Primary endpoint could not be analysed then, but secondary endpoints analyses were available. Over 24 and 36 weeks natalizumab performed better than fingolimod in reducing new T1 Gd+ lesion formation and relapses respectively. No difference was found in the other MRI endpoints (new/enlarging T2 lesions, T2 lesion volumes changes). 50% were naïve patients. A subanalysis found a significant trend for a faster reduction of new T1 Gd+ lesion formation in natalizumab group in the first 8 weeks, which became significant after 12 weeks.

This research confirms again the higher power of natalizumab in controlling inflammatory disease activity with respect to fingolimod. The main limitations of the study are small sample size and short duration of follow-up. However, analyses of MRI endpoints catch the speed of the effect, which was higher for natalizumab.

The paper is well written and clear. However, in order to further improve the work, there are some issues to address.

Statistics
1) Methods, page 10, row 13: Wilcoxon rank-sum test has to be used for dependent sample comparison (i.e. repeated measure within the same group of subjects). Here there is no such analysis, therefore it should be used the Mann-Whitney test for comparison of number of new/newly enlarging T2 lesions and changes in lesion volume between natalizumab and fingolimod groups (two independent samples).

Author response: The Wilcoxon rank-sum test is another name for the Mann-Whitney U test and is used correctly in this instance to compare two independent samples. The Wilcoxon signed-rank test is used for dependent samples.

2) Table 3: the negative binomial model for ARR comparison should also be adjusted for variable "years since first symptoms", which is disease duration, as was done for Cox model for T1 Gd+ lesions.

Author response: The negative binomial model for ARR comparison has been adjusted for years since first symptom, and table 3 (page 12) and figure 2B have been updated with these data.

Others (style, presentation)
3) Introduction, row 22: I would extend this part by citing other important studies, explaining that in

most natalizumab proved to be superior to fingolimod (Carruthers 2014; Kalincik 2015; Barbin 2016; Baroncini 2016; Prosperini 2017), while in some it did not (Braune 2013; Koch-Henriksen 2016).

Author response: The Introduction has been revised to reference these key studies (page 5, paragraph 1).

4) Introduction, row 30 to 42: I would erase all the sentence ("Although…over the duration of the study"), because it is clearly explained in methods section.

Author response: The CONSORT guidelines indicate that specific study objectives should be stated in the Introduction. Thus, these sentences should remain in the Introduction section.

5) Methods, page 8, row 15: I would add "…the sponsor (BIOGEN)…" in the sentence

Author response: The sponsor, Biogen, has been added to this sentence (page 6, paragraph 1).

6) Table 1: standard deviation (SD) should have the same number of decimals of the mean (I would keep only one decimal)

Author response: Table 1 has been updated to provide the same number of decimals for the SD and the mean (page 9).

7) Table 3: I would keep only one decimal for all values (both for means and SD), except for HR and p values.

Author response: Table 3 has been updated to provide the same number of decimals for the SD and the mean (page 12).

8) Table 3: It should be correct with the results of Mann-Whitney test instead of Wilcoxon test

Author response: The Wilcoxon rank-sum test is another name for the Mann-Whitney U test and is used correctly in this instance to compare two independent samples. The Wilcoxon signed-rank test is used for dependent samples.

9) Discussion, page 16, row 10: the same issues as for the introduction in point 3)

Author response: The Discussion has also been updated to reference more recent studies comparing natalizumab and fingolimod (page 14, paragraph 1).

Reviewer: 3
Reviewer Name: Susan Ball
Institution and Country: University of Exeter, UK Please state any competing interests or state 'None declared': None declared

The abstract is clear and easy to understand.
The authors describe the intention-to-treat population as those who were randomised and given at least 1 dose of study drug and who provided any efficacy assessments. Would this be better described as a modified intention-to-treat population?

Author response: The REVEAL study protocol states that this is an intent-to-treat population.

The authors have only been able to conduct analyses of secondary endpoints. Are these analyses based on a smaller sample of participants and/or a shorter follow-up period than originally planned for assessing these endpoints? Were these analyses always intended to be exploratory?

> Author response: The analyses of secondary endpoints are based on a shorter follow-up period than originally planned due to the early closure of the study. For clarity, we have revised the description of the study design in the abstract (page 3, paragraph 2) and the text (page 5, paragraph 2; page 13, paragraph 3) to indicate that these were unplanned exploratory analyses of secondary endpoints.

The statistical methods used seem appropriate and well described. When reporting the results from Negative Binomial regression, would it be possible to report a rate in each trial arm and a rate ratio, with confidence interval? Confidence intervals for differences in rates between the two arms may be more informative than p values.

> Author response: The last row of table 3 presents the ARR for each trial arm and a rate ratio with confidence intervals (page 12).

Table 3 - could a rate ratio be reported in the 'HR (95% CI)' column for ARR on study?

> Author response: A rate ratio for ARR has been added to table 3 (page 12).

Figure 2 - Kaplan Meier plot - check the labeling of vertical axis - seems to be on the wrong scale, compared to the text '22.3%' and '1.9%' on the plot.

> Author response: The vertical axis of figure 2A has been relabelled to indicate percentages.

VERSION 2 – REVIEW

REVIEWER	Damiano Baroncini MS Centre, Galarate Hospital, ASST Valle Olona
REVIEW RETURNED	07-Aug-2020

GENERAL COMMENTS	The authors responded to all my questons. I have no other concern.

REVIEWER	Susan Ball University of Exeter UK
REVIEW RETURNED	16-Aug-2020

GENERAL COMMENTS	Thank you for responding to the reviewer comments. The rationale for the analyses that have been carried out is now clearer, having described it as 'unplanned exploratory analysis of secondary endpoints'. My comment about including an estimated effect (rate ratio) and confidence interval for ARR in Table 3, rather than simply focusing on a p value (which arguably shouldn't be reported at all if these are unplanned exploratory analyses), has been addressed, and provides the reader with more information about the possible difference between the two treatments, and the uncertainty around this estimate. My only further comment is that this rate ratio (~10) is for fingolimod relative to natalizumab (and similarly for other estimated effects in Table 3), whereas in the text the authors

	describe the effects in terms of natalizumab relative to fingolimod. Keeping this consistent throughout may help with understanding. Otherwise I am happy that the authors have addressed the reviewer comments in the revised manuscript.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 2
Reviewer Name: Damiano Baroncini
Institution: MS Centre, Galarate Hospital, ASST Valle Olona

Please state any competing interests or state 'None declared': None declared

1. The authors responded to all my questons. I have no other concern.

Author response: Thank you for your review and for supporting our manuscript.

Reviewer: 3
Reviewer Name: Susan Ball
Institution and Country: University of Exeter, UK

Please state any competing interests or state 'None declared': None declared

1. Thank you for responding to the reviewer comments. The rationale for the analyses that have been carried out is now clearer, having described it as 'unplanned exploratory analysis of secondary endpoints'.

My comment about including an estimated effect (rate ratio) and confidence interval for ARR in Table 3, rather than simply focusing on a p value (which arguably shouldn't be reported at all if these are unplanned exploratory analyses), has been addressed, and provides the reader with more information about the possible difference between the two treatments, and the uncertainty around this estimate. My only further comment is that this rate ratio (~10) is for fingolimod relative to natalizumab (and similarly for other estimated effects in Table 3), whereas in the text the authors describe the effects in terms of natalizumab relative to fingolimod. Keeping this consistent throughout may help with understanding. Otherwise I am happy that the authors have addressed the reviewer comments in the revised manuscript.

Author response: The HRs and rate ratios in Table 3 and Figure 2A have been inverted to correspond with the text, which describes the effects for natalizumab relative to fingolimod (page 13, paragraph 1).